

## ANALIZA TABLICA KONTINGENCIJE

## TABLICA KONTINGENCIJE

- tablica koja u retcima i stupcima sadrži frekvencije atributivnih obilježja
- predstavlja empirijsku razdiobu frekvencija obilježja mjerenih nominalnom ili ordinalnom ljestvicom mjerenja

## TABLICA S "JEDNIM ULAZOM" (1×k)

- opažanja su klasificirana samo po jednom obilježju

### PRIMJER.

	GODINA STUDIJA						UKUPNO
	I	II	III	IV	V	VI	
BROJ STUDENATA	64	48	32	28	18	15	205

## TABLICA S "DVA ULAZA" (r×k)

- opažanja klasificirana po više atributa
- opažanja iz više uzoraka klasificirana po kategorijama jednog atributa

2×2 ....najjednostavnija tablica s "dva ulaza"

obilježje A	obilježje B		UKUPNO
	DA	NE	
DA	$n_{11}$	$n_{12}$	$n_{1y}$
NE	$n_{21}$	$n_{22}$	$n_{2y}$
UKUPNO	$n_{x1}$	$n_{x2}$	$n_{xy}$

- može se promatrati kao:
  - jedan uzorak (sa  $n_{xy}$  ispitanika)
  - dva uzorka (sa  $n_{1y}$ ,  $n_{2y}$  ispitanika)

## TABLICA S "DVA ULAZA" (r x k)

Stručna sprema	Spol		UKUPNO
	Muški	Ženski	
Nezavršena osnovna škola	4	27	31
Osnovna škola	12	35	47
Srednja škola	46	32	78
Viša škola/bakalaureat	12	25	37
Visoka škola/magisterij	52	18	70
Doktorat	11	4	15
UKUPNO	137	141	278

## $\chi^2$ TEST

- ocjena slaganja s poznatom razdiobom
- ocjena razlike razdiobe kategoričkog svojstva u nezavisnim uzorcima
- ocjena razlike dihotomnog svojstva u zavisnim uzorcima

## $\chi^2$ TEST ZA OCJENU SLAGANJA S POZNATOM RAZDIOBOM

- uz unaprijed poznatu razdiobu očekivanih frekvencija, test statistika

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

gdje je:  $O_i$  ..... opažena frekvencija  
 $E_i$  ..... očekivana frekvencija  
 $k$  ..... broj kategorija

ima  $\chi^2$  razdiobu s  
 $df = k - 1 - m$  stupnjeva slobode

$k$  ... broj kategorija  
 $m$  ... broj parametara u modelu koje treba procijeniti

za normalnu razdiobu:  $m = 2$ ;  
 $df = k - 1 - 2 = k - 3$

za binomnu:  $m = 1$   
 $df = k - 1 - 1 = k - 2$

ako je zadana razdioba (ništa ne moramo računati iz podataka):

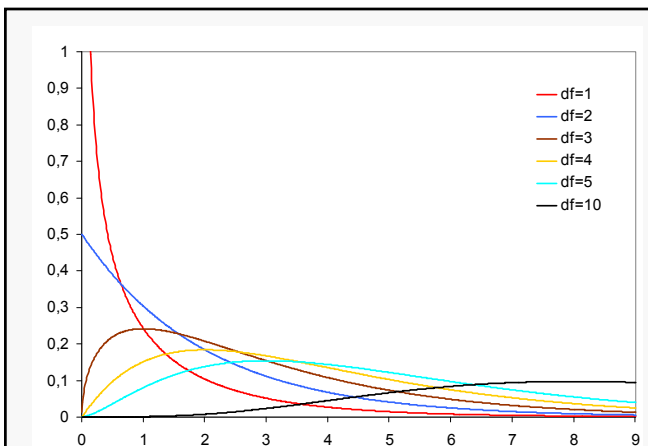
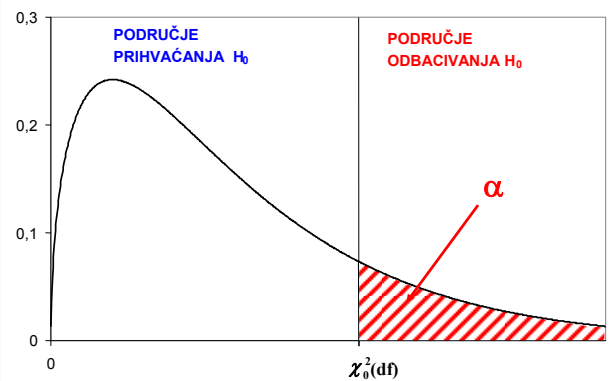
$m = 0$   
 $df = k - 1 - 0 = k - 1$

- UZ  
 $H_0$  ... nema razlike u razdiobi  $O_i$  i  $E_i$

granični  $\chi^2$  za dani  $\alpha$  i  $df$

za  $\chi^2 > \chi^2_{\alpha} \Rightarrow P(\chi^2) < P(\chi^2_{\alpha})$  **ODBACI  $H_0$**

$\chi^2 < \chi^2_{\alpha} \Rightarrow P(\chi^2) > P(\chi^2_{\alpha})$  **PRIHVATI  $H_0$**



Križanjem dviju vrsta biljki dobivena je u sljedećoj generaciji ova razdioba opaženih genotipova:

Genotip	Opažene frekvencije
Aa	53
AA	23
aa	24

Odgovara li ova razdioba očekivanoj razdiobi 2:1:1 uz  $\alpha = 0,01$  ?

genotip	O <sub>i</sub>	E <sub>i</sub>	O <sub>i</sub> - E <sub>i</sub>	(O <sub>i</sub> - E <sub>i</sub> ) <sup>2</sup>	$\frac{(O_i - E_i)^2}{E_i}$
Aa	53	50	3	9	0,18
AA	23	25	-2	4	0,16
aa	24	25	-1	1	0,04
					0,38

$\chi^2 = 0,38$        $df = 3 - 1 = 2$

k \ α	0,01	0,02	0,05
1	6,635	5,412	3,841
2	9,210	7,824	5,991
3	11,345	9,837	7,815

za **df = 2:**  
 $\chi^2_{(\alpha)} = \chi^2_{(0,01)} = 9,210$

$\chi^2 < \chi^2_{(0,01)} \Rightarrow P(\chi^2) > P(\chi^2_{(0,01)})$       **PRIHVATI H<sub>0</sub>**

### Kako dobiti točnu P vrijednost?

- kalkulator vjerojatnosti
  - Excel
  - WEB
  - specijalizirani programi za analizu podataka

### Excel

B4		$\text{fx}$ =CHIDIST(B2;B3)
	A	B
1		
2	Vrijednost $\chi^2$	0.38
3	Broj stupnjeva slobode	2
4	P vrijednost	0.826959

MF Medicinski fakultet Osijek  
Katedra za medicinsku statistiku i medicinsku informatiku

### http://vassarstats.net/

Calculators for Statistical Table Entries

z to P    chi-square to P    t to P    r to P    F to P

Fisher r-to-z transformation    critical values of Q    odds & log odds

For values of degrees of freedom (df) between 1 and 20, inclusive, this unit will calculate the proportion of the relevant sampling distribution that falls to the right of a particular value of chi-square. To proceed, enter the values of chi-square and df in the designated cells and click the «Calculate» button.

Chi-Square	df	P
0.38	2	0.826959

Calculate    Reset

Entered values of df must be between 1 and 20, inclusive.

MF Medicinski fakultet Osijek  
Katedra za medicinsku statistiku i medicinsku informatiku

### Statistica

Probability Distribution Calculator

Distribution: Chi 2

Chi 2: .380000    df: 2

p: .826959

Density Function:

Distribution Function:

MF Medicinski fakultet Osijek  
Katedra za medicinsku statistiku i medicinsku informatiku

### MedCalc – radni list

Data

B3    =CHIDIST(B1;B2)

	A	B
1	Chi	0.38
2	Df	2
3	P	0.826959134
4		
5		
6		

MF Medicinski fakultet Osijek  
Katedra za medicinsku statistiku i medicinsku informatiku

## Izveštavanje rezultata

"Nije uočena razlika između razdiobe opaženih genotipova i očekivane razdiobe 2:1:1 ( $\chi^2$  test,  $P = 0,827$ )."

ili

"Nije uočena razlika između razdiobe opaženih genotipova i očekivane razdiobe 2:1:1 ( $\chi^2 = 0,380$ , 2 stupnja slobode,  $P = 0,827$ )."

## $\chi^2$ TEST ZA NEZAVISNE UZORKE

postupak:

- formirati tablicu kontingencije ( $r \times k$ )
- na osnovu postavljene hipoteze izračunati očekivane frekvencije
- test statistika dana je sa:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

gdje je:  $r$  .....broj redaka  
 $k$  .....broj stupaca

ima  $\chi^2$  razdiobu  $s$

$df = (r - 1) \cdot (k - 1)$  stupnjeva slobode

### VAŽNE NAPOMENE:

- u tablicu smijemo unijeti **SAMO APSOLUTNE FREKVENCIJE**
- uzorci moraju biti **nezavisni**
- u **2 x 2** tablici:
  - **NITI JEDNA** očekivana frekvencija **ne smije biti < 5**
  - Yatesova korekcija (umanjiti svaku razliku  $O - E$  prije kvadriranja za 0,5) ili "N-1" Hi-kvadrat test (modificirana kratka formula)
- ako je u  $r \times k$  tablici  $E < 5$  u više od 20 % polja, **NE MOŽEMO KORISTITI  $\chi^2$  TEST**  
rješenje:
  - spajanje susjednih razreda (frekvencija susjednih polja)
  - Fisherov egzaktni test



Pri istraživanju djelovanja nekog cjepiva, opažena je sljedeća učestalost oboljenja kod određene grupe ljudi:

	Cijepljeni	Necijepljeni	Ukupno
Oboljeli	5	10	15
Nisu oboljeli	159	117	276
Ukupno	164	127	291

Postoji li povezanost između učestalosti bolesti i cijepljenja (je li učestalost bolesti različita kod cijepljenih i necijepljenih) uz  $\alpha = 0,01$  ?

$H_0$  .... učestalost je ista kod cijepljenih i necijepljenih

iz  $H_0 \Rightarrow$  proporcije oboljelih trebaju biti jednake u obje skupine

zajednička proporcija oboljelih:  $z_{po} = \frac{15}{291} = 0,051546$

ukupno oboljeli

sveukupno ispitanika

zajednička proporcija zdravih:  $z_{pz} = \frac{276}{291} = 0,948454$

ukupno zdravi

sveukupno ispitanika

E oboljelih: ukupno cijepljeni

u grupi cijepljenih.....  $164 \cdot 0,051546 = 8,45$

u grupi necijepljenih..  $127 \cdot 0,051546 = 6,55$

ukupno necijepljeni

zajednička proporcija oboljelih

E zdravih: ukupno cijepljeni

u grupi cijepljenih.....  $164 \cdot 0,948454 = 155,55$

u grupi necijepljenih..  $127 \cdot 0,948454 = 120,45$

ukupno necijepljeni

zajednička proporcija zdravih

	Cijepljeni	Necijepljeni	Ukupno
Oboljeli	5 (8,45)	10 (6,55)	15
Nisu oboljeli	159 (155,55)	117 (120,45)	276
Ukupno	164	127	291

\*

O <sub>i</sub>	E <sub>i</sub>	O <sub>i</sub> - E <sub>i</sub>	(O <sub>i</sub> -E <sub>i</sub> ) <sup>2</sup>	(O <sub>i</sub> -E <sub>i</sub> ) <sup>2</sup> /E <sub>i</sub>
5	8,45	-3,45	11,9025	1,41
10	6,55	3,45	11,9025	1,82
159	155,55	3,45	11,9025	0,08
117	120,45	-3,45	11,9025	0,10
				$\chi^2 = 3,41$

**Yates-ova korekcija:**

O <sub>i</sub>	E <sub>i</sub>	(O <sub>i</sub> - E <sub>i</sub> ) <sub>corr</sub>	(O <sub>i</sub> -E <sub>i</sub> ) <sub>corr</sub> <sup>2</sup>	(O <sub>i</sub> -E <sub>i</sub> ) <sub>corr</sub> <sup>2</sup> /E <sub>i</sub>
5	8,45	-2,95	8,7025	1,03
10	6,55	2,95	8,7025	1,33
159	155,55	2,95	8,7025	0,06
117	120,45	-2,95	8,7025	0,07
				$\chi^2 = 2,49$

$\chi^2 = 2,49$   
za  $\alpha = 0,01$  i  $df = 1$ :  $\chi_0^2 = 6,635$   
 $\chi^2 < \chi_0^2 \Rightarrow P > 0,01$   
 $\Rightarrow$  ne postoji povezanost između učestalosti bolesti i cijepljenja

$\alpha$	0,01	0,02	0,05
1	6,635	5,412	3,841
2	9,210	7,824	5,991
3	11,345	9,837	7,815

$\chi^2$  RAZDIOBA

**točna P vrijednost.....**

B4		$f_x = \text{=CHIDIST}(B2;B3)$
A	B	
1		
2	Vrijednost $\chi^2$	3.41
3	Broj stupnjeva slobode	1
4	P vrijednost	0.064802

**bez korekcije**

B4		$f_x = \text{=CHIDIST}(B2;B3)$
A	B	
1		
2	Vrijednost $\chi^2$	2.49
3	Broj stupnjeva slobode	1
4	P vrijednost	0.114572

**s korekcijom**

**2 x 2 tablice – kratka formula za  $\chi^2$  test**  
- za tablicu 2 x 2 sa sljedećim oznakama:

Obilježje A	Obilježje B		UKUPNO
	0	1	
0	a	b	m
1	c	d	n
UKUPNO	r	s	N

$$\chi^2 = \frac{N(ad - bc)^2}{mnr}$$

**"N - 1"  $\chi^2$  test**

Obilježje A	Obilježje B		UKUPNO
	0	1	
0	a	b	m
1	c	d	n
UKUPNO	r	s	N

$$"N - 1" \chi^2 = \frac{(N - 1)(ad - bc)^2}{mnr}$$

$\chi^2$  test (kratka formula) i "N - 1"  $\chi^2$  test za podatke iz primjera

	Cijepljeni	Necijepljeni	Ukupno
Oboljeli	5 (a)	10 (b)	15 (m)
Nisu oboljeli	159 (c)	117 (d)	276 (n)
Ukupno	164 (r)	127 (s)	291 (N)

$$\chi^2 = \frac{N(ad-bc)^2}{mnr s} = \frac{291 \cdot (5 \cdot 117 - 10 \cdot 159)^2}{15 \cdot 276 \cdot 164 \cdot 127} = \frac{293917275}{86227920} = 3,409$$

$$\begin{aligned} \text{"N-1" } \chi^2 &= \frac{(N-1)(ad-bc)^2}{mnr s} = \frac{290 \cdot (5 \cdot 117 - 10 \cdot 159)^2}{15 \cdot 276 \cdot 164 \cdot 127} = \\ &= \frac{292907250}{86227920} = 3,397 \end{aligned}$$

točna P vrijednost.....

B4		=CHIDIST(B2;B3)
A	B	
1		
2	Vrijednost $\chi^2$	3.409
3	Broj stupnjeva slobode	1
4	P vrijednost	0.064842

kratka formula

B4		=CHIDIST(B2;B3)
A	B	
1		
2	Vrijednost $\chi^2$	3.397
3	Broj stupnjeva slobode	1
4	P vrijednost	0.065315

"N - 1"  $\chi^2$

MedCalc – za tablice 2 x 2 koristi "N - 1"  $\chi^2$  test

Izveštavanje rezultata

"Nije uočena povezanost učestalosti bolesti i cijepljenja ( $\chi^2$  test, P = 0,065)."

ili

"Nije uočena razlika razdiobe obolijevanja u skupinama cijepljenih i necijepljenih ispitanika ( $\chi^2$  test, P = 0,065)."

Izveštavanje rezultata

"Nije uočena povezanost učestalosti bolesti i cijepljenja ( $\chi^2 = 3,397$ , 1 stupanj slobode, P = 0,065)."

ili

"Nije uočena razlika razdiobe obolijevanja u skupinama cijepljenih i necijepljenih ispitanika ( $\chi^2 = 3,397$ , 1 stupanj slobode, P = 0,065)."

$\chi^2$  TEST ZA ZAVISNE UZORKE (McNemarov test)

- testiranje značajnosti razlike (ili vjerojatnosti povezanosti) između podataka dobivenih na uzorcima parova

$$\chi^2 = \frac{(|b-c| - 1)^2}{b+c}$$

Yatesova korekcija

b, c .... frekvencije parova koji se ne slažu po prisutnosti obilježja

OBILJEŽJE A		UZORAK I	
		DA	NE
UZORAK II	DA	a	b
	NE	c	d

Skupina od 75 bolesnika praćena je tijekom 20 godina. Ispitanici su s obzirom na težinu svrstani u skupinu normalne i prekomjerne (overweight) težine.

TEŽINA		Nakon 20 godina		Ukupno
		Normalna	Prekomjerna	
Na početku	Normalna	26	14	40
	Prekomjerna	3	32	35
Ukupno		29	46	75

Je li se težina ispitanika promijenila tijekom promatranog perioda uz  $\alpha = 0,05$  ?

### McNemarov test – pomoću tablica

$$\chi^2 = \frac{(|14 - 3| - 1)^2}{14 + 3} = \frac{10^2}{17} = 5,882$$

$$df = 1, \alpha = 0,05$$

$$\chi_0^2 = 3,841$$

$$\chi^2 > \chi_0^2 \Rightarrow P < 0,05$$

k \ α	0,01	0,02	0,05
	1	6,635	5,412
2	9,210	7,824	5,991
3	11,345	9,837	7,815

točna P vrijednost.....

- za  $\chi^2 = 5,882$  i 1 stupanj slobode:

B4		fx = =CHIDIST(B2;B3)	
	A	B	
1			
2	Vrijednost $\chi^2$	5.822	
3	Broj stupnjeva slobode	1	
4	P vrijednost	0.015827	

### MedCalc – McNemarov test (binomni egzaktni test)

McNemar test (paired proportions)

2x2 table

	Pos.	Neg.	
Pos.	26	14	53.3%
Neg.	3	32	46.7%
	38.7%	61.3%	

Results

Difference	14.67%
95% CI	2.98% to 20.94%
Exact probability	P = 0.0127

Comment:

Buttons: Clear, Test, Exit

### Izveštavanje rezultata

"Težina ispitanika se promijenila tijekom promatranog perioda (McNemarov test,  $P = 0,013$ )."

ili

"Težina ispitanika se promijenila tijekom promatranog perioda (razlika proporcija 14,7 %, 95 % raspon pouzdanosti od 3,0 % do 20,9 %, McNemarov test,  $P = 0,013$ )."